

1. 서론

본 연구는 다양한 배치 및 샘플링 시간으로 특징지어지는 배치 공정 데이터를 위한 견고한 머신러닝 모델을 구축할 수 있도록 동적 시간 왜곡(DTW)과 유전자 알고리즘(GA)을 결합하는 방법을 제안합니다. 이 종합적인 모델은 초기 공정 조건, 공정 변수의 시계열 데이터, 제품 품질 등 주요 요소를 통합하며, 목표 제품 품질 값을 달성하기 위해 초기 공정 조건과 배치 프로파일로 정의되는 공정 변수의 시계열 데이터를 동시에 설계할 수 있습니다. 중합체 생산을 위한 실제 반배치 공정 데이터를 활용하여 제안된 방법이 배치 공정 중 및 배치 종료 시점에서 제품 품질을 정확하게 예측할 수 있음을 확인하였고, 동시에 합리적인 초기 공정 조건 및 배치 프로파일을 설계할 수 있음을 입증하였습니다.

2. 방법

2.1 동적 시간 왜곡(DTW)

본 연구에 사용된 데이터셋은 연속적인 공정 변수 값을 포함한 시계열 데이터로, 배치 공정에서 샘플링 시간과 배치 시간의 차이로 인해 각 시계열 데이터의 길이가 다릅니다. 따라서 배치 공정에서 샘플 간의 유사성을 계산하기 위해 DTW를 도입하였습니다. DTW는 두 시계열 데이터 집합 간의 유사성을 계산합니다. 예를 들어, 시계열 데이터 T 를 (t_1, t_2, \dots, t_p) 라 하고, 다른 시계열 데이터 U 를 (u_1, u_2, \dots, u_q) 로 표시합니다. 각 요소 (i) 와 (j) 간의 거리 $(\delta(i, j) = |t_i - u_j|)$ 로 정의하며, T 와 U 간의 유사성 $(S(T, U))$ 는 다음과 같이 계산됩니다:

$$S(T, U) = \min_w \sum_{k=1}^r \delta(w_k)$$

여기서 (w) 는 최소 거리 조합을, (r) 은 조합 수를 나타냅니다. 본 연구에서는 dtwalign 도구를 사용하여 DTW를 계산하였습니다.

2.2 DTW와 GA를 활용한 종료 지점 예측 및 배치 프로파일 설계

훈련 데이터의 특징은 배치 공정의 초기 공정 조건, 배치 시간, 공정 변수의 시계열 데이터, 종료 지점의 특성으로 나누어집니다. 종료 지점의 특성은 y 로 설정되며, 배치 공정의 초기 공정 조건은 x 로 설정됩니다. 공정 변수의 시계열 데이터에 대해 모든 샘플 간 각 공정 변수에 대한 DTW를 계산하였으며, 훈련 데이터의 각 샘플과 공정 변수의 DTW 값을 x 에 추가하였습니다. 초기 공정 조건의 수를 (k) , 공정 변수의 수를 (m) , 훈련 데이터의 샘플 수를 (n) 이라 할 때, x 변수의 수는 $(k + mn)$ 이 됩니다. 회귀 분석을 통해 x 와 y 간 머신러닝 모델 $(y = f(x))$ 을 구축하였습니다.

테스트 데이터 또는 새로운 데이터의 경우 초기 공정 조건과 훈련 데이터와의 DTW를 시계열 데이터에서 계산하여 y 값을 예측하기 위한 x 로 입력합니다. 또한 목표 y 값을 얻기 위해 초기 공정 조건과 공정 변수의 시계열 데이터, 즉 배치 프로파일을 제안하였습니다. 유전자 알고리즘(GA)을 사용하여 초기 조건과 시계열 데이터를 포함하는 데이터를 생성하고, 예측된 y 값이 목표 y 값과 일치하도록 반복하였습니다.

3. 결과 및 논의

3.1 결과 개요

평균값(AV)과 최종값(FV), 그리고 가장 짧은 배치 시간(ST)을 조정하여 각 시계열 데이터를 준비하였고, 이 방법을 제안된 DTW 기반 방법과 비교하였습니다. 부분 최소 제곱 회귀(PLS), 릿지 회귀, 절대값 축소 및 선택 연산자(LASSO), 엘라스틱넷, 서포트 벡터 회귀(SVR), 랜덤 포레스트, 가우시안 프로세스 회귀(GPR)를 사용하여 회귀 분석을 수행하였으며, 각 시계열 데이터 처리 방식에 대해 예측 능력이 가장 우수한 방법을 선택하였습니다. 본 연구에서는 배치 수가 적기 때문에 이중 교차 검증(DCV)을 통해 예측 정확도를 평가하였습니다.

DCV 예측 결과는 표 1과 표 2 및 그림 1과 그림 2에 나타나 있습니다. 결정 계수 (r^2)와 예측의 평균 제곱근 오차(RMSE)가 각 표에 제공됩니다. 표 1에 따르면 (y_1)에 대한 예측 정확도는 AV 방법에서 가장 높았으나, 그림 1에서 네 가지 방법 모두에 대해 예측 결과에 큰 차이가 없음을 확인할 수 있었습니다. (y_1)은 배치 시간 및 공정 변수의 배치 프로필에 크게 의존하지 않고 초기 공정 조건에 의해 주로 결정됩니다. 반면, (y_2)의 경우 DTW 방법이 가장 높은 (r^2) 값을 나타냈으며, 그림 2에서 볼 수 있듯이 큰 (y_2) 값을 정확하게 예측한 것은 DTW 방법뿐이었습니다. (y_2)는 초기 공정 조건뿐만 아니라 배치 프로필에도 영향을 받는 것으로 확인되었습니다. 본 연구에서 제안된 DTW 방법이 배치 중 및 종료 시점에서 물리적 특성을 정확하게 예측할 수 있음을 확인하였습니다.

유전자 알고리즘(GA)을 기반으로 한 역 분석을 통해 구축된 (y_1)과 (y_2) 예측 모델을 사용하여 목표 값을 각각 50과 1523으로 설정하였습니다. 첫 번째 단계의 시간은 305분이었고, 두 번째 단계의 후보 시간은 30분에서 380분까지 설정하였습니다. GA는 두 번째 단계의 시간, 초기 공정 조건, (T_r), (T_j), (F)의 배치 프로필을 최적화하여 예측된 y 값과 목표 y 값 사이의 거리를 줄였습니다.

4. 결론

본 연구에서는 배치 시간이 다양한 배치 공정에서 제품 특성 예측 모델을 DTW와 머신러닝을 활용해 구축하는 방법을 제안하였으며, 유전자 알고리즘(GA)을 이용하여 목표 품질 값을 달성하기 위한 초기 공정 조건, 배치 시간, 배치 프로필을 설계하는 방법을 제시하였습니다. 시계열 데이터의 샘플 간 유사성을 각 공정 변수별로 DTW를 통해 계산하였고, 이를 배치 공정의 특징인 x 변수로 사용하였습니다. 이로써 제품 품질의 목표 값을 예측하고 이를 달성하기 위해 초기 공정 조건, 배치 시간 및 배치 프로필을 GA를 통해 설계할 수 있게 되었습니다. 실제 반배치 공정 데이터셋을 활용한 검증 결과, 본 연구에서 제안한 방법이 배치 공정에서 중합체 재료의 두 가지 특성을 정확하게 예측하고, 기존 데이터셋에 없는 품질 사양을 만족하는 초기 공정 조건, 배치 시간 및 배치 프로필을 설계할 수 있음을 확인하였습니다.

본 연구에서 제안한 방법은 배치 공정의 제품 품질을 예측하는 머신러닝 모델을 구축하고, 모델의 역 분석을 통해 배치 공정을 최적화하여 운영 효율성을 향상시키고 불량 제품을 줄이는 데 기여할 수 있습니다. 또한, 본 연구에서 제안된 방법을 통해 공정 변수의 시간 변화를 예측할 수 있어 온라인 배치 공정 최적화 및 실시간 피드백 제어가 가능해집니다. 향후 연구에서는 공정 변수의 미래 시간 변화 예측에 기반한 공정 제어 연구를 포함할 필요가 있습니다.